DOCUMENT RESUME

ED 044 049 95 EM 008 543

AUTHOR Ferguson, Douglas, Ed.
TITLE Information Retrieval (SPIRES) and Library Automation (BALLOTS) at Stanford University.
INSTITUTION Stanford Univ., Calif. Computation Center.
SPONS AGENCY National Center for Educational Research and Development (DHEW/CE), Washington, D.C.; National Science Foundation, Washington, D.C.
PUB DATE Nov 70
GRANT OEG-1-7-071145-4428(095)
NOTE 16p.

EDRS PRICE EDRS Price MF-$0.25 HC-$0.90
DESCRIPTORS Bibliographic Citations, Book Catalogs, Information Centers, Information Dissemination, Information Needs, *Information Retrieval, Information Services, *Library Acquisition, Library Extension, Library Reference Services, Library Research, Library Technical Processes, Research Libraries, *University Libraries
IDENTIFIERS BALLOTS, Bibliographic Automation Large Library Operations, SLAC, SPIRES, Stanford Linear Accelerator Center, *Stanford Physics Information Retrieval System

ABSTRACT

At Stanford University, two major projects have been involved jointly in library automation and information retrieval since 1968: BALLOTS (Bibliographic Automation of Large Library Operations) and SPIRES (Stanford Physics Information Retrieval System). In early 1969, two prototype applications were activated using the jointly developed systems software: an acquisition system in the Main Library (BALLOTS I) and a bibliographic retrieval system (SPIRES I) in the Stanford Linear Accelerator Center (SLAC) Library. The goals for BALLOTS II are: responsiveness to library users, efficient operation, generality, performance monitoring, and flexibility for future improvement. For SPIRES II, the goals are related to five areas: data source and content, cost and customers, search facilities, feedback, and record modification. The work of SPIRES and BALLOTS has potential beyond its immediate applications. It can support socially significant research, as in the fields of ecology and urban studies; and with remote terminals, it can provide information quickly and at the sit e of research. Such a comprehensive information facility would be, in a sense, an "extended library;" and receiving daily use, it could be offered at a favorable cost-benefit ratio. (MF)

BALLOTS

INFORMATION RETRIEVAL (SPIRES)

AND

LIBRARY AUTOMATION (BALLOTS)

AT

STANFORD UNIVERSITY

NOVEMBER, 1970

I/F-70-1

ACKNOWLEDGMENT

The activity reported here was performed under:

Grant Nos. OEG-1-7-071145-4428 (095) and OEG-0-70-5237 with the Office of Education, United States Department of Health, Education and Welfare (BALLOTS--Bibliographic Automation of Large Library Operations Using a Time-Sharing System), and

Grant Nos. GN 600, GN 742, GN 830, and GN 830.1 with the National Science Foundation (SPIRES--Stanford Physics Information REtrieval System).

Grantees undertaking such projects are encouraged to express freely their professional judgement in the conduct of the project. Points of view or opinions stated do not, therefore, necessarily represent official agency position or policy.

TABLE OF CONTENTS

Preface iii

The Problem Context 1

SPIRES/BALLOTS Project 1

BALLOTS I and SPIRES I 2

BALLOTS II and SPIRES II: System Development 3

BALLOTS II and SPIRES II: Goals 5

Toward an Information Facility 7

BIBLIOGRAPHY: PUBLICATIONS AND REPORTS 10

PREFACE

This report is the result of contributions and comments from several people. John Schroeder, SPIRES/BALLOTS Manager for Technical Development, contributed material on the system development process from which the System Development section was written. The entire text was read and criticized by the following: Professor Edwin B. Parker, Principal Investigator for SPIRES; Mr. David C. Weber, Director of Libraries; Mr. Allen B. Veaner, Principal Investigator for BALLOTS; and Mr. Hank Epstein, SPIRES/BALLOTS Project Director. Miss Jennifer Hartzell of the SPIRES/BALLOTS Documentation Office made many helpful editorial suggestions.

Douglas Ferguson, Editor
SPIRES/BALLOTS Project
Stanford, University

INFORMATION RETRIEVAL AND LIBRARY AUTOMATION: SPIRES/BALLOTS

THE PROBLEM CONTEXT

The publication explosion, a compelling need for access to information, and rapid library growth are not unique to Stanford University. At Stanford, a commitment has been made to deal with the information problems of the university by improving library service and developing a campus based bibliographic retrieval system. Using the tools of computing technology and library systems analysis, computer specialists, librarians, and behaviorial scientists have joined in exploring the bibliographic requirements of a major university community and creating new systems to meet those requirements.

Library automation requires a major effort in system development and sizeable expenditures for computer equipment. Computerized information storage and retrieval requires a similar investment in hardware and software. Both undertakings have common conceptual problems in such areas as bibliographic file organization and on-line searching. Each derives concrete benefits from the other: bibliographic files created in the process of library automation are available for generalized retrieval uses, and complex retrieval routines are available for searching library bibliographic files.

SPIRES/BALLOTS PROJECT

At Stanford, two major projects have been involved jointly in library automation and information retrieval since 1968. One is BALLOTS (Bibliographic Automation of Large Library Operations Using a Time-Sharing System), funded by the Office of Education; the other is SPIRES (Stanford Physics Information REtrieval System--informally known as the Stanford Public Information REtrieval System), funded by the National Science Foundation. The purpose of this collaboration is to create the common software required to support both the BALLOTS and SPIRES applications. The joint effort is overseen by the SPIRES/BALLOTS Executive Committee, chaired by Professor William F. Miller, Vice President for Research. Professor Edwin B. Parker of Stanford's Institute for Communication Research is Principal Investigator for SPIRES. Mr. Allen R. Veaner, Assistant Director of University Libraries for Bibliographic Operations, is Principal Investigator for BALLOTS. Management responsibility for the joint project has been delegated to the Stanford Computation Center. Mr. Hank Epstein, Project Director, holds a joint appointment with the Computation Center and the University Libraries.

The Stanford project structure and system development philosophy reflect both the common uses and individual needs of BALLOTS and SPIRES. The concept of shared facilities results in system software and hardware designed to service both the BALLOTS and SPIRES applications. Two examples of common software are an on-line text editor and file- handling and task-scheduling routines. These shared facilities can service bibliographic input and specialized research files. Examples of shared hardware facilities are a central processing unit and direct access devices (making possible shared files). Combining resources in this system development effort reduces the cost of creating common facilities and provides a pool of skilled manpower resources for each project.

BALLOTS I AND SPIRES I

In 1967 the Stanford University Libraries and the Stanford Institute for Communication Research began research projects with funds from the Office of Education (BALLOTS) and the National Science Foundation (SPIRES). In 1968 the two projects came under the policy direction of the SPIRES/BALLOTS Executive Committee, thus formalizing the shared perspective and close collaboration of the projects.

Stanford University was an appropriate setting in which to initiate research and development in bibliographic retrieval. Strong interest in automation was felt in all areas of the Stanford University Libraries and especially by its Associate Director (now Director), David C. Weber, and Assistant Director for Bibliographic Operations, Allen B. Veaner. During the period from 1964 to 1966, the library had achieved a remarkably successful computer-produced book catalog for the J. Henry Meyer (Undergraduate) Library. Professor Edwin B. Parker and his colleagues at the Institute for Communication Research even then were applying to computer systems the behavioral science analysis that they and others had applied already to print, film, and television media. The Stanford Campus Facility had an IBM 360 model 67 computer, a locally developed time-sharing system, and a first-rate programming staff associated with one of the nation's leading computer science departments. A close working relationship between the University Libraries, the Computation Center, and the Institute for Communication Research formed a firm foundation for research and development.

The combined project software development group applied themselves to writing programs necessary for bibliographic retrieval. In the library, an analysis and design group worked closely with the library staff in studying library processes and defining requirements. This joint effort created a prototype system that could be used in the Main Library and by Stanford faculty and students, primarily high- energy physicists.

In early 1969, two prototype applications were activated using the jointly developed systems software: an acquisition system was established in the Main Library (BALLOTS I) and a bibliographic retrieval system (SPIRES I) was established for a group of high-energy physicists.

Centralized management of library input was handled by two newly created departments, Data Preparation and Data Control. Several terminals were installed in the Main Library for on-line searching, and a terminal was placed in the Physics Library. An on-line In Process File was created, consisting of 30 percent of the roman-alphabet acquisition material ordered by the library. A specially trained staff performed on-line searching daily during regular library hours. This prototype acquisition system operated during most of 1969, demonstrating the technical feasibility of the combined project goals. It was studied and evaluated by the library systems and programming staffs, who reviewed the human, economic, and technical requirements of a library bibliographic retrieval system.

At the Stanford Linear Accelerator Center (SLAC) Library, a file of preprints in high-energy physics was created through SPIRES I. This file is still active; records of new preprints are added weekly, and a note is made of any preprint that is published. Input is via an IBM 2741 typewriter terminal in the SLAC Library. Regular library staff at SLAC handle inputs and updating. Searching can be done by author, title, date, and citation. The preprint file contains approximately 6,500 documents, including all the high-energy- physics preprints received in the SLAC Library from March 1968 to the present. "Preprints in Particles and Fields," a weekly listing of preprints, also is produced from SPIRES I. After an initial period of support by the Division of Particles and Fields of the American Physical Society, it now is supported partially by subscriptions.

BALLOTS II AND SPIRES II: SYSTEM DEVELOPMENT

The results of operating the prototype applications (BALLOTS I and SPIRES I) were encouraging, particularly with respect to the advantages and practicality of utilizing common software. Under actual operating conditions, feasibility and usefulness were estab- lished and a wealth of knowledge was gained. Joining the library and retrieval application areas served by shared facilities (hardware and software) was shown to be a rewarding approach.

BALLOTS I and SPIRES I resulted from a development process in which user requirements were analyzed, programs written and tested, and prototypes created and evaluated. Librarians, behaviorial scientists, library systems analysts, and computer specialists collaborated over an extended period of time. This development process was a major accomplishment. It made possible the definition of a production bibliographic retrieval system with distinctive hardware and software requirements.

The creation of a production system for library automation (BALLOTS II) and for generalized information storage and retrieval (SPIRES II) requires continuing a comprehensive system development process. Within the framework of this process tasks are defined, assigned, and coordinated. The system development process for the creation of BALLOTS II and SPIRES II has six phases:

Phase A: Preliminary Analysis
Phase B: Detailed Analysis
Phase C: General Design
Phase D: Detailed Design
Phase E: Implementation
Phase F: Installation

Preliminary Analysis involved defining goals, describing the user environment, analyzing the existing system, selecting the system scope, and establishing the gross technical feasibility of the selected first implementation scope. These factors are described in detail in a System Scope Document that was the main output of the Preliminary Analysis Phase.

Detailed Analysis enumerates minutely the requirements that the manual-automated system must meet. In it (1) performance requirements are stated quantitatively, including response time, hours of on-line accessibility, allowable mean failure time, maximum allowable recovery time, and similar factors. (2) Record input and output are determined in terms of volume, growth, and fluctuations. Timing considerations for batch input and output are determined, in order to plan for scheduling requirements. (3) All input and output record, screen, and document formats are determined character by character. (4) Rules transforming input data elements into output data elements are formulated and tabulated. (5) The upper bounds of development and operating costs are established.

General Design encompasses both system externals (procedures, training, reorganization, etc.) and system internals (alternative hardware and software solutions to the stated requirements). As a result an overall software-hardware configuration is selected and outlined in a General Design Document.

Detailed Design completes the internal and external design, creates implementation and testing plans, and provides programming specifications. These factors are incorporated in a Detailed Design Document.

In the Implementation Phase, user documentation is created and personnel training begins. Programs are coded and checked. Testing is carried out and the results are evaluated. Programs, maintenance documentation, and test reports are prepared in this phase.

In the Installation Phase, training of all personnel is completed, files are converted, and, after the automated system has operated parallel to the manual system for a time, the changeover is made to the automated system. Performance statistics then are collected and a support plan and project history are written.

Each phase description necessarily has been abbreviated. Not all activities or outputs have been described. Activities in some phases overlap and feed back to redefine previous activities. In the instal- lation phase, a "Wishbook" that has been maintained through all the phases is put in final form. The Wishbook is very important because it is a link to successive development iterations. It contains information on the capabilities, services, and operational character- istics whose desirability became apparent during the development process but which could not be included owing to time, cost, or technical constraints. The Wishbook also contains information on any internal (programming or hardware) and external (user or procedural) operational deficiencies that are determined after the system has been running for some time. These findings will be considered in designing new portions of the system and will aid in improving its overall design.

This outline of the system development process guides SPIRES/ BALLOTS II development from the definition of goals to the instal- lation of a fully operational system.

BALLOTS II AND SPIRES II: GOALS

The project goals for library automation (BALLOTS), generalized information storage and retrieval (SPIRES), and shared facilities are interrelated. The goals of shared facilities (hardware and software) support and serve the goals of BALLOTS and SPIRES.

BALLOTS

As the major information center of a large academic institution, the library must respond rapidly, effectively, and economically to the university community. The library is a complex combination of people, machines, and records that organize and open up the major bibliographic resources of the university to students and faculty. It reflects the needs and priorities of a changing university environment. The university library is also part of a larger network of information sources that includes other research libraries, the Library of Congress, and specialized information storage agencies.

The essential goals of BALLOTS are found in a library system (both manual and automated sections) which is: USER RESPONSIVE. The system adapts to the changing bibliographic requirements of diverse user groups within the university community. COST

COMPETITIVE. The system provides fast, efficient internal processing of increasing volumes of processing transactions. GENERALIZABLE. The system is not just an attempt to automate portions of the existing manual system. It is based on the actual operating requirements of library processing; it is not completely dependent on the existing procedural, organizational, or physical setting. PERFORMANCE ORIENTED. The system provides the library and university administration with data which are useful for measuring internal processing performance and user satisfaction. FLEXIBLE. The system has the capability to expand in order to embrace a broader range of services and a wider group of users. Service via terminals will be available throughout the campus. The system will be able to link up with and serve other information systems and use effectively national data sources.

These goals will be expressed in specific capabilities which will, among other things, minimize manual filing; eliminate many clerical tasks now performed by professionals; increase self-service efficiency; and provide mechanisms for recording the user's suggestions. The effect of these computer capabilities will be to reduce drastically errors associated with manual sorting, typing, and hand transcription; to speed the flow of material through library processing; to aid book selection by providing fast access to central machine files; and to enable librarians to advise a patron of the exact status of a work about which he is inquiring. In summary, responsiveness to library users, efficient operation, generality, performance monitoring, and flexibility for future improvement are the essential goals of library automation.

SPIRES

The SPIRES generalized information storage and retrieval system will support the research and teaching activities of the library, faculty, students, and staff. Each user will be able to define his requirements in a way that automatically tailors the system response to fit his individual needs. The creation of such a system is a major activity involving the study of users, source data, record structure, and file organization, as well as considerable experimentation with facilities. The SPIRES system will be charac- terized by flexibility, generality, and ease of use. SPIRES goals are related to five specific areas: DATA SOURCE AND CONTENT. A generalized information storage and retrieval capability will store bibliographic, scientific, administrative, and other records in machine-readable form. Collections will range from large public files, converted from centrally produced machine-readable data, to medium-small files, created from user generated input (faculty and student files). SEARCH FACILITIES. The system will provide the capability for searching files interactively (on-line), via a computer terminal; on a batch basis, by grouping requests and submitting them on a regular schedule; and on a standing request basis, in which a search query is routinely passed against certain files at specified intervals. FEEDBACK. Reports will be provided on how

frequently various system elements are used. These will include statistical analyses of user difficulties and system errors. RECORD MODIFICATION. Update and edit capability will be provided on a batch basis or on-line; options for update will come at the level of record, data element, and character string within a data element. COST AND CUSTOMERS. The cost of the services provided should be sufficiently low to enable a wide range of customers to cost justify their use of the system. The variety of services should be sufficiently great to encourage a growing body of users. A range of services at various cost levels must be offered to permit users to select the type of service which meets their needs within their financial limitations.

BALLOTS AND SPIRES SHARED FACILITIES

Shared facilities are hardware and software designed to provide concurrent service to BALLOTS and SPIRES applications. Since the sharing of such resources represents a substantial saving to all applications served, maximum attention will be given to the sharing concept. Whenever possible, advantage will be taken of the economies gained through providing major facilities for multiple applications. HARDWARE. The hardware environment will provide reliable, economical, and flexible support of the applications it encompasses. SOFTWARE. The software, which will consist of an operating system, an on-line executive program, a terminal handler, a text editor, and many other facilities, will be used jointly by various applications. GENERALITY/EXPANDABILITY. The shared facilities will be designed to allow current applications to be expanded as well as to allow new applications to be added without modifying previous ones.

TOWARD AN INFORMATION FACILITY

The work of SPIRES and BALLOTS has potential beyond bibliographic retrieval and library automation applications. There is a growing need for computer and other information retrieval services in support of socially significant research. Such research is being conducted in the developing fields of ecology and urban studies, to name two areas.

Several capabilities and services are required. Data banks of bibliographic and other information are needed for studies that draw upon several disciplines. Strong disciplinary information systems (e.g., psychology) and centralized national systems (e.g., Educational Resources Information Center--ERIC) produce large amounts of data on magnetic tapes. In addition, at Stanford, data are generated in on-campus research and at nearby centers. An infor- mation facility with large-scale storage equipment and sophisticated program capabilities could create and maintain data banks derived from such input sources.

Data selected from large machine-readable files could be subjected to further computer processing. Programs that perform mathematical or statistical analysis could be used to produce evidence for a problem solution that may not have been considered when the data were first gathered. Fresh insights often can be obtained without the necessity of generating large amounts of new data. Similarly, new data could be added to an existing file and up-to-date analyses performed to confirm or extend the conclusions of previous studies.

In addition to the large amounts of information now in machine- readable form, even larger amounts are now available in microforms. These include microfiche, microcards, and microfilm. Computer- generated on-line indexes to massive microform files are a form of information retrieval that an information facility could provide.

In social research there is an increasing premium placed on providing information quickly and at the site of research--which is often beyond the university campus. A computer information facility can meet these time and distance requirements. Remote terminals in a nearby city can access a Stanford computer. Direct-access storage devices operating with time-sharing software can provide immediate interactive response. Stanford has several years' experience in operating a multi-user interactive system with a network of users in the San Francisco Bay Area.

An information facility based on SPIRES and BALLOTS would combine production operating characteristics and sophisticated search and retrieval capabilities. Reliability, security, fast recovery, and cost acceptability are required to support library and administrative operations. Ease of use for people with nontechnical backgrounds (faculty or students) and extensive search request capabilities are needed by the SPIRES users. Through a combined facility, librarians would be able to use one or more search programs and a researcher would be able to access library files. Both of these activities could be carried out simultaneously from different locations. Common software such as a terminal handler would serve all user groups.

In a sense, a comprehensive information facility is an "extended library." The boundaries of this library are not physical walls but the telecommunication limits of the facility terminals. A researcher will not need to go to the library catalog to search for material--a "catalog" will be as close as a terminal.

Such an information facility would be used by several major user groups. These groups include not only librarians and university administrative personnel but also researchers at Stanford and in nearby locations. In the past, and even now, no one of these user groups could afford to have a computer facility devoted to its own information needs. As an information facility

is designed to serve the daily operational needs of the university and the special information needs of various research groups, computerized information services can be offered at a favorable cost-benefit ratio.

BIBLIOGRAPHY: PUBLICATIONS AND REPORTS

<NOTE: ED numbered documents are available from ERIC Document Reproduction Service, 4936 Fairmont Avenue, Bethesda, Maryland 20014.
PB numbered documents are available from Clearinghouse for Federal Scientific and Technical Information, Springfield, Virginia 22151.>

AUTOMATION NEWSLETTER: PROJECT BALLOTS. Automation Division, Stanford University Libraries. No. 1, June 1969; No. 2, September 1969; No. 3-4 (combined issue), March 1970. <Available without charge from SPIRES/BALLOTS Documentation Office, Cypress Hall Annex, Stanford, California 94305.>

BALLOTS PROGRESS REPORT 3/27/69 - 6/26/69. SPIRES/BALLOTS Project, Stanford University, Stanford, California. 402 pp. ERIC document number 030 777; microfiche $1.50, hard copy $20.20.

Parker, Edwin B. "Behaviorial Research in the Development of a Computer-Based Information System." Delivered at the Conference on Communication among Scientists and Technologists: The Study of the Production, Dissemination and Use of Information by Scientists and Technologists, October 28-30, 1969, at The Johns Hopkins University, Baltimore, Maryland. To be published in the Conference Proceedings.

Parker, Edwin B. "Democracy and Information Processing." EDUCOM (Bulletin of the Inter-University Communications Council) 5, No. 4(1970):2-6.

Parker, Edwin B. "Developing a Campus Based Information Retrieval System." In PROCEEDINGS, STANFORD CONFERENCE ON COLLABORATIVE LIBRARY SYSTEMS DEVELOPMENT (Stanford University, Stanford, California, October 4-5, 1968), Stanford University Libraries, Stanford, California, 1969, pp. 213-30. ERIC document number ED 031 281; microfiche $1.00, hard copy $11.50.

Parker, Edwin B. "Information Utilities and Mass Communication." In Nie, N., and H. Sackman (eds.), INFORMATION UTILITIES AND SOCIAL CHOICE. Montvale, New Jersey: AFIPS Press, 1970.

Parker, Edwin B. "Potential Interrelationships Between Library and Other Mass Media Systems." Delivered at the Conference on Interlibrary Communications and Information Networks, September 28 - October 2, 1970, at Warrenton, Virginia.

Parker, Edwin B. "A System Theory Analysis of Feedback Mechanisms for Information Systems." Delivered at the FID International Congress of Documentation, September 21-24, 1970, at Buenos Aires, Argentina. To be published in the Congress PROCEEDINGS.

RECON Working Task Force (Allen B. Veaner, Member). CONVERSION OF RETROSPECTIVE CATALOG RECORDS TO MACHINE-READABLE FORM. Library of Congress, Washington D.C., 1969.

RECON Working Task Force (Allen B. Veaner, Member). "Levels of Machine-Readable Records." JOURNAL OF LIBRARY AUTOMATION 3, No. 2 (1970):122-27.

SPIRES ANNUAL REPORTS
1967 Report, 57 pp. ERIC document number ED 617 294; microfiche $.50, hard copy $3.05.

1968 Report, 107 pp. Clearinghouse document number PB-184 960; microfiche $.65, hard copy $3.00.

1969-70 Report, 129 pp. ERIC document number ED 04 602. <The 1969-70 Report is available without charge from SPIRES/BALLOTS Documentation Office, Cypress Annex, Stanford, California 94305.>

SPIRES REFERENCE MANUAL. SPIRES/BALLOTS Project, Stanford University, Stanford, California. 61 pp. January 1969. Revised July 1969.

SPIRES/BALLOTS REPORT. A 15-minute, color, 16mm film giving an overview of the library automation and information retrieval problem in general and of Stanford's approach to it. Written and directed by D.B. Jones. Produced by the Department of Communications, Stanford University, 1969. <Copies may be rented from Extension Media Center, University of California, Berkeley 94720. Rental charge, $15.00 for 24 hours; purchase price, $180.00.>

SYSTEM SCOPE FOR LIBRARY AUTOMATION AND GENERALIZED INFORMATION STORAGE AND RETRIEVAL AT STANFORD UNIVERSITY. SPIRES/BALLOTS Project, Stanford University, Stanford, California. February 1970. Second printing, April 1970. 157 pp. ERIC document number ED 038 153. <Available from the SPIRES/BALLOTS Documentation Office, Cypress Hall Annex, Stanford, California 94305, for $7.50 prepaid.>

Veaner, Allen B. "The Application of Computers to Library Technical Processing." COLLEGE AND RESEARCH LIBRARIES 31, No. 2 (1970):36-42.

Veaner, Allen B. "Major Decision Points in Library Automation." Delivered at the 75th Annual Meeting of the Association of Research Libraries, January 17-18, 1970, at Chicago, Illinois. COLLEGE AND RESEARCH LIBRARIES 31, No. 5 (1970):299-312. The full text as delivered appears in the published meeting minutes, pp. 2-33.

Veaner, Allen B. "Stanford University Libraries, Project BALLOTS, A Summary." In PROCEEDINGS, STANFORD CONFERENCE ON COLLABORATIVE LIBRARY SYSTEMS DEVELOPMENT (Stanford University, Stanford, California, October 4-5, 1968), Stanford University Libraries, Stanford, California, 1969, pp. 42-49. ERIC document number ED 031 281; microfiche $1.00, hard copy $11.50.

Weber, David C. "Personnel Aspects of Library Automation." Delivered at the Information Science and Automation Division Program Meeting of The American Library Association, June 28 - July 3, 1970, at Detroit, Michigan.